



INTRO TO DATA SCIENCE

LECTURE 18: WHERE TO GO NEXT?

Jason Dolatshahi
Data Scientist, EveryScreen Media

LAST TIME(S):

- NETWORK GRAPHS**
- CAUSALITY (SEAN TAYLOR)**
- RECOMMENDER SYSTEMS @ SPOTIFY**

QUESTIONS?

I. COURSE RECAP

II. READ STUFF

III. DO STUFF

INTRO TO DATA SCIENCE

I. COURSE RECAP

	<i>continuous</i>	<i>categorical</i>
<i>supervised</i>	regression	classification
<i>unsupervised</i>	dimension reduction	clustering

	<i>continuous</i>	<i>categorical</i>
<i>supervised</i>	regression	classification
<i>unsupervised</i>	dimension reduction	clustering

regression: linear & basis function regression

classification: knn, naïve Bayes, logistic regression, decision trees, support vector machines

clustering: k-means

dimension reduction: pca, svd, nonlinear methods

recommender systems: content-based & collaborative filtering

ensemble methods: bagging, boosting, random forests

network graphs: centrality and diffusion

dev tools: UNIX, R, Python

database tools: MySQL, Mongo DB

map-reduce: big data & horizontal scaling, Hadoop

iterative problem solving

feature spaces

supervised vs unsupervised learning

loss functions (eg, solving ML problems via optimization)

regularization

bias/variance tradeoff

relational db's vs nosql

curse of dimensionality

THOUGHT EXPERIMENT

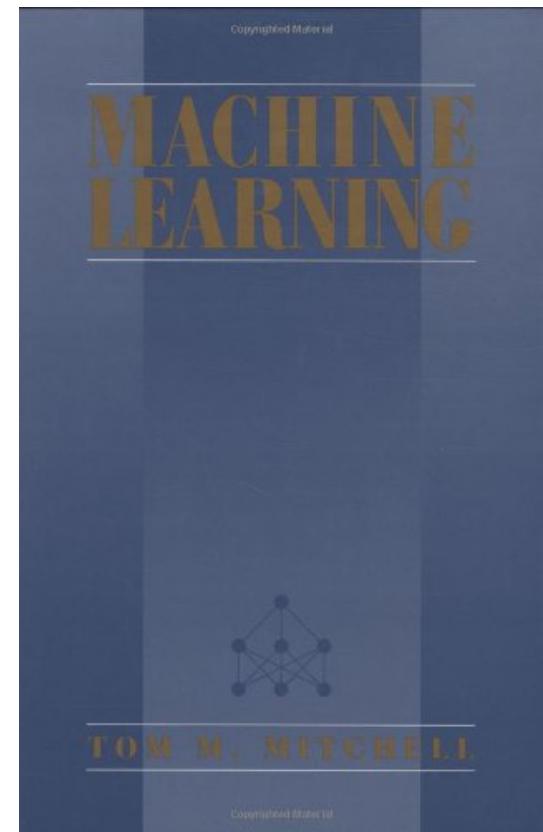
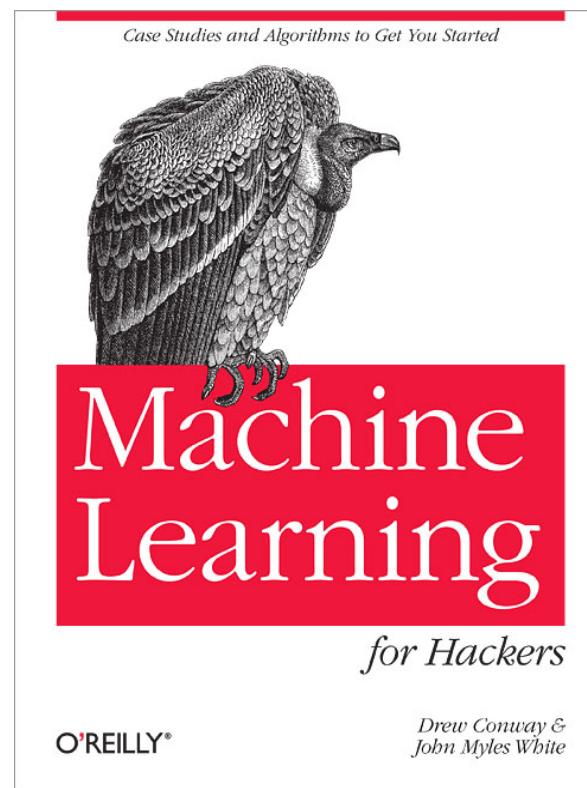
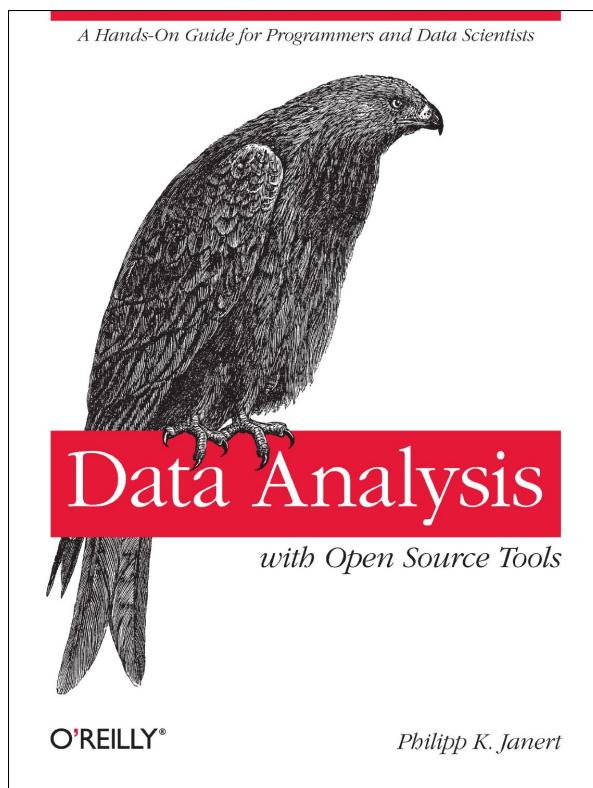
Think about how you would discuss your understanding of these concepts (eg, in an interview).

INTRO TO DATA SCIENCE

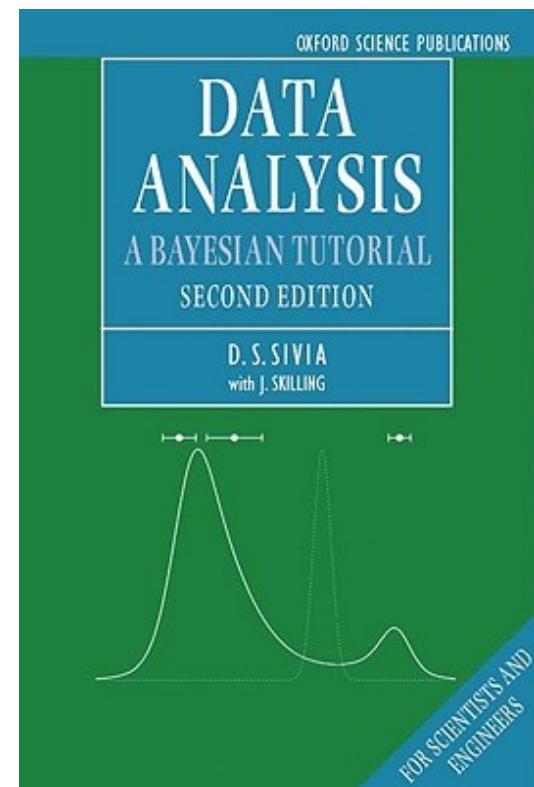
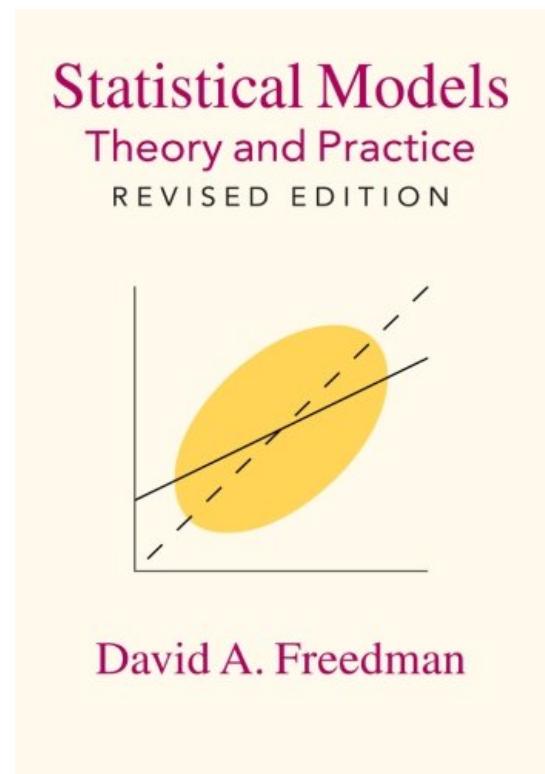
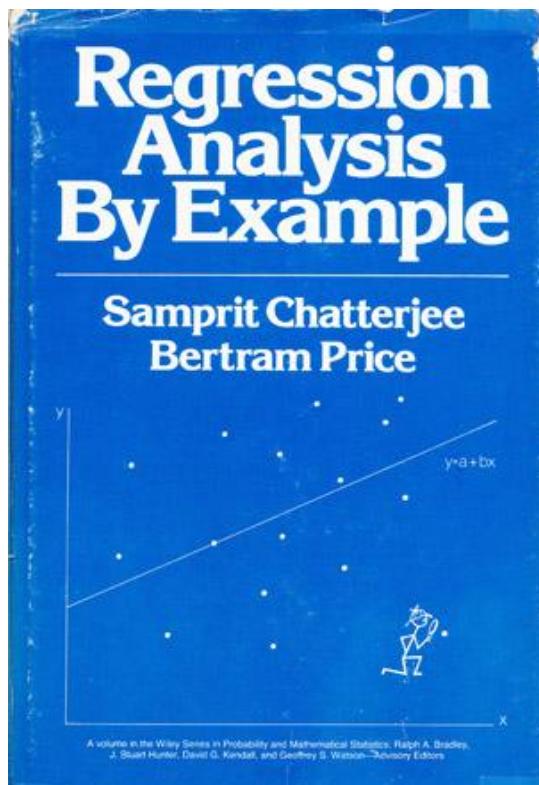
II. READ STUFF

BOOKS - INTRO

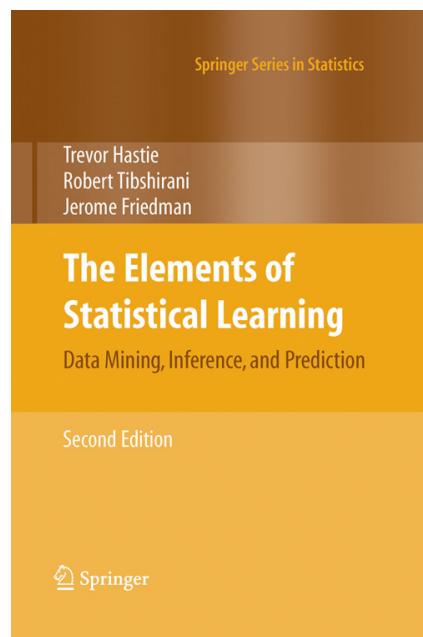
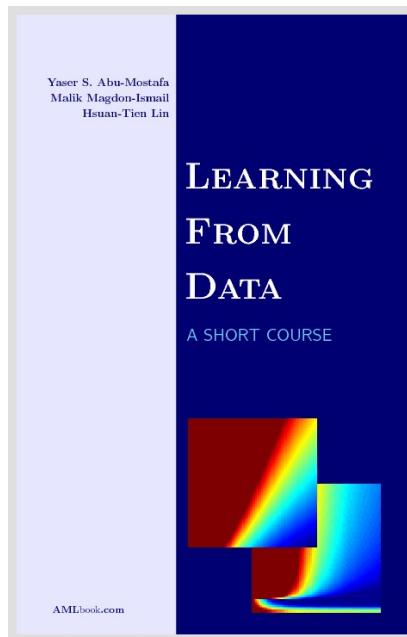
11



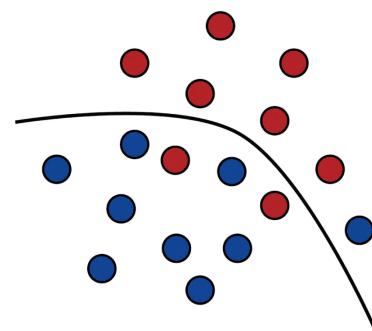
BOOKS – PARAMETRIC METHODS



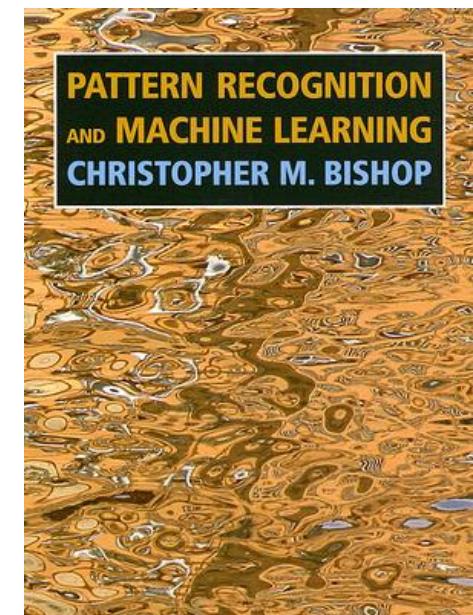
BOOKS – ML THEORY



Foundations of
Machine Learning

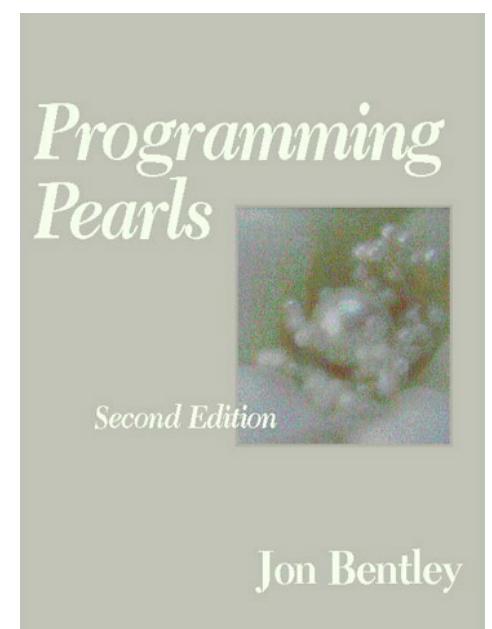
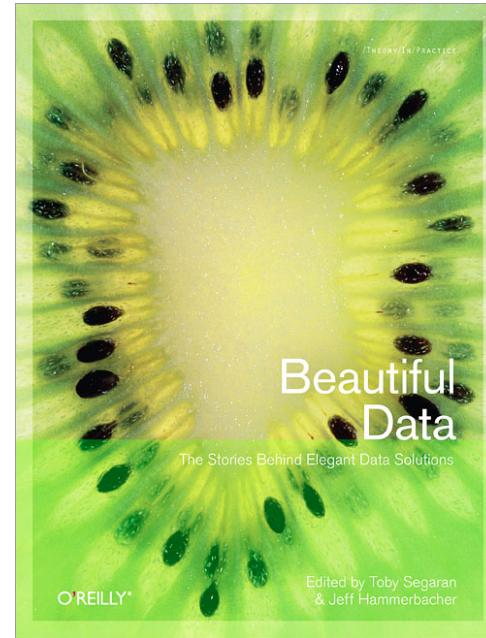
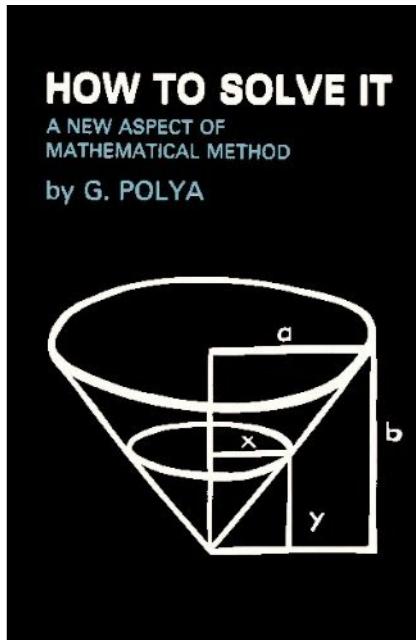
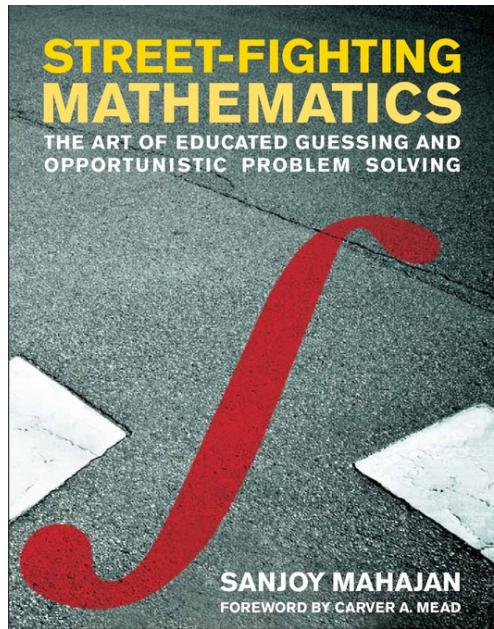


Mehryar Mohri,
Afshin Rostamizadeh,
and Ameet Talwalkar



BOOKS – PROBLEM SOLVING

14



<https://news.ycombinator.com/>

<http://www.r-bloggers.com/>

<http://www.johnmyleswhite.com/>

<http://www.hilarymason.com/>

<http://blog.echen.me/>

<http://andrewgelman.com/>

<http://hunch.net/>

<http://conductrics.com/data-science-resources/>

<http://fivethirtyeight.blogs.nytimes.com/>

- <https://www.coursera.org/course/ml> Stanford ML course
- <http://www.youtube.com/playlist?list=PLD63A284B7615313A> Caltech ML course
- http://videolectures.net/Top/Computer_Science/Machine_Learning/
- <http://www.hilarymason.com/tag/video/>
- <http://ocw.mit.edu/courses/mathematics/18-06-linear-algebra-spring-2010/video-lectures/>
- <http://harvarddatascience.com/>
- <http://www.autonlab.org/tutorials/> (Andrew Moore tutorials page)
- <http://www.columbia.edu/~lah2178/index/Courses.html> (Lauren Hannah courses)

INTRO TO DATA SCIENCE

II. DO STUFF

GO TO MEETUPS!

18

- <http://www.meetup.com/ny-tech/> (NY Tech meetup)
- <http://www.meetup.com/nycpython/>
- <http://www.meetup.com/NYC-Predictive-Analytics/>
- <http://www.meetup.com/nyhackr/> (formerly R meetup)
- <http://www.meetup.com/NYC-Data-Engineering/>
- <http://www.meetup.com/hackabit/> (Bitly hackathons)
- <http://www.meetup.com/hackshackersny/>
- <http://www.meetup.com/Hadoop-NYC/>
- <http://www.meetup.com/NYC-Machine-Learning/>
- <http://www.meetup.com/Prince-Building-Tech-Talks/>

Project Euler.net



DataKind
USING DATA IN THE SERVICE OF HUMANITY

kaggle

what subjects interest you?

read stuff

do stuff

share your results!

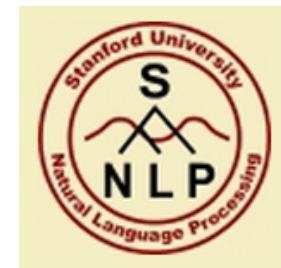
what subjects interest you?

read stuff

do stuff

share your results!

natural language processing



or



?



DEV SKILLS

22

low level languages:

C/C++, Java

scripting languages:

JavaScript, Scala, PHP

niche languages:

Haskell, Go, Erlang

viz languages:

d3, Processing

jason@everyscreenmedia.com

jason.dolatshahi@gmail.com

INTRO TO DATA SCIENCE

THANKS!